

Neural Networks and Spectral Feature Selection for Retrieval of Hot Gases Temperature Profiles

Esteban García-Cuesta
Departamento de Física
egc@fis.uc3m.es

Inés M. Galván
Departamento de Informática
igalvan@inf.uc3m.es

Antonio J. de Castro
Departamento de Física
decastro@fis.uc3m.es

Universidad Carlos III Avda. Universidad 30
28911 Leganés(Madrid), España

Abstract

Neural networks appear to be a promising tool to solve the so-called inverse problems focused to obtain a retrieval of certain physical properties related to the radiative transference of energy. In this paper the capability of neural networks to retrieve the temperature profile in a combustion environment is proposed. Temperature profile retrieval will be obtained from the measurement of the spectral distribution of energy radiated by the hot gases (combustion products) at wavelengths corresponding to the infrared region. High spectral resolution is usually needed to gain a certain accuracy in the retrieval process. However, this great amount of information makes mandatory a reduction of the dimensionality of the problem. In this sense a careful selection of wavelengths in the spectrum must be performed. With this purpose Principal Component Analysis technique is used to automatically determine those wavelengths in the spectrum that carry relevant information on temperature distribution. A multilayer perceptron will be trained with the different energies associated to the selected wavelengths. The results presented show that multilayer perceptron combined with Principal Component Analysis is a suitable alternative in this field.

1. Introduction

Progress in optoelectronic technologies during last decade has led to the fabrication of new sensors to measure the radiated energy focused on new measurement concept based on high spectral resolution measurements. High resolution measurements implies a better understanding of the physical properties related to the radiated energy. However the amount of information increases in a way that makes difficult the use of conventional data regression techniques to retrieve the physical information involved in the problem.

Many problems could be found in these kind of retrievals, related with data dimension or complexity of regression models which implies a large number of operations to solve it. This large amount of data led to try new techniques for the retrieval of this information.

The artificial neural networks seem to be an interesting alternative technique to solve this kind of problems. One important advantage of neural networks in this field is their speed. Once the neural network has been trained the inversion method is almost instantaneous in comparison to regressions models. Another advantages over classical physical-statistical techniques are that do not need a good initial condition for the inversion and do not need a rapid direct model for iterative inversion algorithms.

The goal of this study is to present an inversion method that retrieves the temperature profile of a hot gas cloud composed by CO₂ and water vapor from spectroradiometric measurements. This problem is related to ill-posed problems and corresponds with inverse radiative problems[13]. To do it, a multilayer perceptron (MLP) approximation has been adopted.

In previous works, neural networks have been used to retrieve the atmospheric temperature[3], although the problem here is quite different because of the temperature intervals are bigger and hotter, and in atmospheric retrieval the length is known a priori which simplifies the inverse model. In our case we have a complex relationship between optical path and temperature whose influence vary in the straightforward model following Beer's law, and in the inverse model we do not know how this relationship is because both factors have a non-linear influence in all the spectrum measurements.

In a first approach, MLP has been trained using the whole spectrum which implies a large number of input nodes in the network with redundant and noisy information. Usually as more data are used better results would be expected in the retrieval algorithm, but at the same time a

problem of data management begin. The high dimensionality of input space could prevent an appropriate performance of MLP and an elevated time learning might be required. To reduce the dimensionality of the input space, a second approach is proposed. In this approach, a reduction dimensionality technique is used to decide in an automatic way the most relevant wavelengths in the spectrum. After, MLP is trained using those wavelengths with the purpose to reduce as the complexity of the network as the cost for the learning phase.

In this work, Principal Component Analysis (PCA) has been used to reduce the number of input neurons of MLP. PCA is a multivariate statistical analysis introduced by Pearson (1901)[14], and developed independently by Hotelling (1933)[8]. PCA involves a mathematical procedure that transforms a number of correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

The main intention using PCA technique, as previously mentioned, is to reduce the dimensionality finding a new input dimension in which the correlated variables are eliminated and preserve most of the variation about temperature and length. Thus, it will be possible to approach the problem with MLP with high resolution measurements and evaluate whether the performance of the MLP when the input variables are those corresponding with the results of data set reduction is higher or like equal minimum than the performance of MLP trained with the whole spectrum.

The rest of the paper is organized as follows: In section 2 a context description of retrieval of temperature profiles in flames is made. Section 3 describes the simulations carried out and the obtained results. This section includes also the use of PCA to reduce the dimensionality of the input space for the MLP. Finally, conclusions are presented in section 4.

2 Description of the Problem: Retrieval of Temperature Profiles in Flames

In a industrial fuel fired furnace, it is very important to have devices that monitor and control the combustion process in order to minimize pollutant emissions as well as to optimize energy losses. Flame temperature appears, among others, as a very important parameter to be monitored[15][18][12][11]. Conventional temperature monitoring devices such thermocouples are intrusive and disturb the measurement, and they must undergo the harsh furnace environment. Remote optical measurements are more suitable because they are non-intrusive. Ultraviolet, visible and infrared detectors have been used in flame monitoring systems[15]. Infrared sensing appears to be very

promising, because the hot gases in the flame, mainly carbon dioxide (CO_2) and water vapor (H_2O) exhibit important emission bands in the infrared region. A recent trend in flame thermometry is based on spectrometric measurements that discriminate the received energy as function of the wavelength. An example is the application of the so-called emission-transmission method by using tunable infrared laser and optical fiber[12]. This technique is an active technique, because it uses an infrared source in addition to the sensor system. These methods are very sensitive, but their high cost and complexity makes them not very suitable for routine operations in industrial furnaces.

This paper presents some results within the framework of an authors general proposal to use passive infrared spectroscopy to recover the temperature profile inside a hot gas cloud composed by CO_2 and water vapor, representative of a fossil fuel combustion. The experimental equipment to be used is a sensor (for instance a commercial spectroradiometer) that measures directly the spectral distribution of radiated energy by the flame in the infrared spectral range. In this way the infrared source is not necessary (passive measurement). The selected spectral range is 3-5 μm , because carbon dioxide presents a strong emission band in this region[4]. Moreover, this range is commonly implemented in infrared systems, because atmospheric absorption is not very important¹.

The problem to obtain the temperature profile from such a spectrum is not straightforward. Energy emission at each wavenumber depends in a non-linear way on parameters like the spatial distribution of temperature and gas concentrations, or the gas cloud width. Moreover each wavenumber emission depends in a different way on these parameters. To illustrate these difficulties, figure 1 shows a theoretical calculation of the CO_2 emission band for three different situations: a CO_2 cloud 100 cm wide at 1000 K, the same cloud but 1 cm wide and a CO_2 cloud 100 cm wide with a temperature exponentially decayed from 1000 K to 400 K. It is very clear from the figure that the shape of the temperature profile and the width of the cloud are key parameters to understand the information involved in the energy emission spectrum.

3 Temperature Retrieval Using Neural Networks

Neural networks techniques can be used as an approach to solving problems of fitting experimental data. In the context of approximation of nonlinear maps, the architecture most widely used is the MLP. It is relatively quite straight-

¹Note the magnitude at the X axis: it is common for spectroscopists to use the wavenumber ν instead of the wavelength λ . Wavenumber is defined as the inverse of wavelength, and it is typically measured in cm^{-1} in the infrared spectral region.

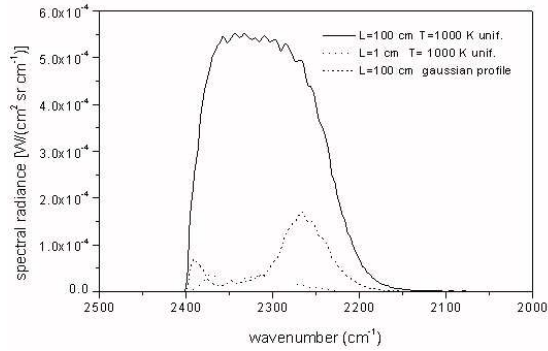


Figure 1. CO₂ emission band calculated for: a) a cell of width $L=100$ cm and uniform temperature $T=1000$ K; b) a cell of width $L=1$ cm and uniform temperature $T=1000$ K; c) a cell of width $L=100$ cm and a temperature distribution (Gaussian profile).

forward to use and it has been proven by different authors that they are universal approximators ([5], [7]), in the sense that any continuous function may be represented by a MLP with one hidden layer.

The MLP neural network has their neurons divided into disjoint subsets, called layers. It is composed by the input layer, that receives the external data, the last layer that acts as the output of the network and the intermediate layers called hidden layers. Two consecutive layers are connected via synaptic connections that have associated a real number, called weight of the connection. Each neuron in a layer forms a weighted sum of the inputs from previous layers to which it is connected, adds a threshold value and apply the activation function (sigmoidal function) producing its activation or output value. The adjustable parameters of the MLP (weights and thresholds) are determined using the well known backpropagation learning algorithm [17], which updates the parameters to minimize the error between the vector output of the network and the desired vector output.

In the next, the procedure to obtain the data sets to train MLPs is explained. After, the results of different experimental simulations are shown. In a first approach, MLPs use as input variables the whole spectrum. The second approach try to reduce the input dimensionality of the network using the PCAs technique.

The performance of different approaches is measured, in one hand, in terms of mean square error over the training and test data sets and, in other hand, in terms of average temperature error per profile and for the hottest cell. The hottest cell is used as criterion because it's retrieval is the most difficult due to the fact that energy emitted by this cell is absorbed by the others which behave as a mask.

3.1 Experimental Data Sets

The data set is composed of large number of synthetic emission spectra generated with a computer code developed at University Carlos III (CASIMIR)[6] based on the well known HITRAN/HITEMP[16] spectral database. The total number of cases simulated are 1040 covering many possible sceneries of a typical flame combustion. Data set generation has been performed under the following assumptions:

- Synthetic spectra will correspond to energy emission of hot gas cloud of width L . Temperature and gas concentrations present gradients inside the cloud.
- The spectral range selected for this data set is 2000cm^{-1} - 2500cm^{-1} . Most of the commercial infrared instruments have capabilities to measure in this range. In this spectral range, the CO₂ emission band is by far the most important emission feature, being the water emission nearly negligible. Due to this fact, only the emission associated to the CO₂ will be considered.
- For retrieval scheme, we have used an spatial discretization in a basic case with five cells of equal width ($L/5$). Each cell has an average value of temperature and gas concentration.
- The objective of this study is focus on the dependence of the spectral energy distribution on temperature profile and length. For this reason the concentration profiles for carbon dioxide and water vapor will keep unchanged for the whole data set. Numerical values for these concentrations have been selected from typical combustion experiments.
- Four basic temperature profiles have been chosen to simulate different temperature gradient. The step between the temperature of two consecutive synthetics flames is $\Delta T = 50$ K, with a variation in the hottest cell among 540 K and 1140 K. And for each of these variations of temperature, a variation of cell's length is done. These variation have an step of $\Delta w = 0.02$ meters for each cell which means a total step variation $\Delta W = 0.1$ meters, covering a range between 0.1 and 2 meters. These profiles adjusted to a spatial discretization of five cells can be seen in figure 2. In these simulations the most unfavorable situation from energy detection point of view has been always chosen: the cell corresponding to the hottest part of the cloud is located opposite from the location of the sensor system. All the value ranges for temperature and length have been chosen to be representative for hot gases clouds associated to fossil fuel combustion.
- Experimental noise of spectra has not been simulated in order to avoid the extraction of features associated

with it when principal component analysis is being used.

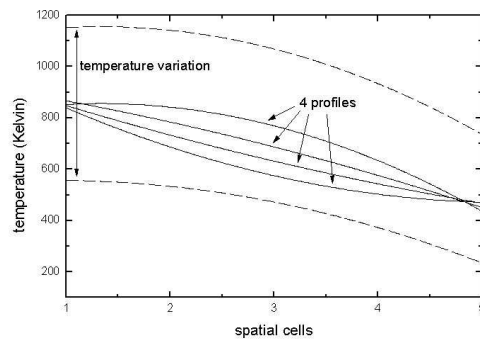


Figure 2. Temperature profiles variations for retrieval scheme.

3.2 First Approach: Using the Whole Frequency Spectrum

In a first phase, MLP has been trained using the whole spectrum. In this case the experiments have done with 247 dimensions (medium spectral resolution) but experiments with higher spectral resolution -1,000 dimensions in the same spectral region and near 10,000 if many regions are selected- will be done in the future. As it has been explained in subsection 3.1 the profiles of temperature are discretized to five cells, so the output layer will have six neurons, one for each temperature cell and another one for the total length². We must include the length because of both parameters, temperature and optical depth or total length, have influence in the composition of the spectrum in agreement to the Lambert-Beer law. Different architectures of MLP varying the number of hidden neurons have been trained until to reach the minimum value in validation error, not allowing overfitting. In table 1, the mean square error over the trained and test data for different architectures are shown. The table also included the mean error per profile and the mean error on the hottest cell.

It is possible to observe that for a small number of hidden neurons the MLP converges without good results due to the high number of inputs, so more hidden neurons are needed. For an architecture of one level with 20 or 30 hidden neurons better approximations are obtained with a relative error of 0.8% in the hottest cell, a mean temperature error of 7.44 K and a relative error of length of 2.6% which is an acceptable accuracy in our study.

²We are assuming in the discretization that all the cells have the same length so we do not include one per each cell.

Table 1. Errors for the hot gas temperature retrieval using a MLP with 247 inputs and different architectures.

Hidden neurons	MSE Train	MSE Test	Mean error per profile (K)	Mean error hottest cell (K)
10	0.00430	0.00800	22.41	13.60
10x10	0.00244	0.00230	13.48	7.47
20	0.00064	0.00112	7.66	6.57
40	0.00043	0.00102	7.88	7.82
30	0.00026	0.00101	7.44	6.63

3.3 Second Approach: Feature Selection for Dimension Reduction with PCA

In order to introduce a priori information in the MLP helping the learning process and speed it up, a dimension reduction approach has been assumed. This reduction try to conserve all the information of any possible scenery in a few components stressing the importance of that wavelengths whose influence in the temperature and length profile are important. This reduction is known as feature selection and the technique used to make it, is the widely known PCA. The central idea of PCA is to reduce the dimensionality of data set in which there are a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This reduction is achieved by transforming to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables[10]. The new base is composed of a set of axes which will be orthogonal between them and are calculated as a lineal combination of the old base.

Also a dimension reduction could be done using the projections of the original data over this new base, but during the experiments realized the results obtained have been always worst than with a feature selection method and consequently has been rejected.

Let $C = \{e_1^M, \dots, e_n^M\}$ be a data set of n spectrum of dimension $M = 247$ variables. Let Σ be the covariance matrix of the data set C with dimension $M \times M$. Let V the $M \times M$ matrix with columns equal to eigenvectors of Σ and let L be the diagonal $M \times M$ matrix with the M associated eigenvalues (by definition $\Sigma \cdot V = V \cdot L$).

The selection of m specific channels from M variables where $m \ll M$, allows to work with lower dimensionality. This m subset of variables contains virtually all the information available in M variables. The problem then is to find the value of m , and to decide which subset or subsets of m variables are best. Here we want to find that variables which best represent the internal variation of C to find out which channels are significant (feature selection). In other cases the linear correlation between PCs and channels are used to interpret the physical meaning[9], or to get a first retrieval approximation[2]. To resolve the question about

Table 2. Cumulated percentage of variance for spectrum data set generated.

Number of PCA components	Cumulated variance %
1	88.58
2	97.05
3	99.46
4	99.81
5	99.88
6	99.94
7	99.97
8	99.98
9	99.98
10	99.98
11	99.98
12	99.99

how many m variables we have to consider, we will check the number of PCs that account for most of the variation in a spectrum e_x of the data set C . This can also be interpreted as finding the effective dimensionality of e_x . If e_x can be successfully described by only m PCs, then it will often be true that p can be replaced by a subset m (or perhaps slightly more) variables, with a relative small loss of information[10][136-137].

First step in a PCA approach is to calculate the covariance matrix of dimensions 247×247 . The eigenvalues and the corresponding eigenvectors of this covariance matrix are computed using the QR and QL algorithms[19]. The results obtained as cumulative percentage of variance are shown in table 2. Between fifth and sixth principal component, around the 99.9% of the total variation it is covered and the spectrum could be reconstruct almost without error. Furthermore we have visualized the projections of the data set for this first principal components trying to find clusters and we have found that with five PCs we can do a first approximation clustering (k-means) by temperature and total length. It means that the projections in these first five PCs have information about temperature and length scales. To select of subset of variables from this first five PCs a peak peaks algorithm has been chosen trying to search for the most important groups coefficients of each eigenspectrum³ in absolute value.

The results can be seen in figure 3 and the selected data is mark with arrows and their wavelength. Because we are not trying to get the best reduction, but to improve or compare the retrieval results with the use of all the wavelengths, a first approximation of $m = 17$ has been adopted.

As in the previous phase, different number of hidden

³Eigenspectrum is the eigenvector matrix which corresponds to the Σ matrix of the spectrum data set.

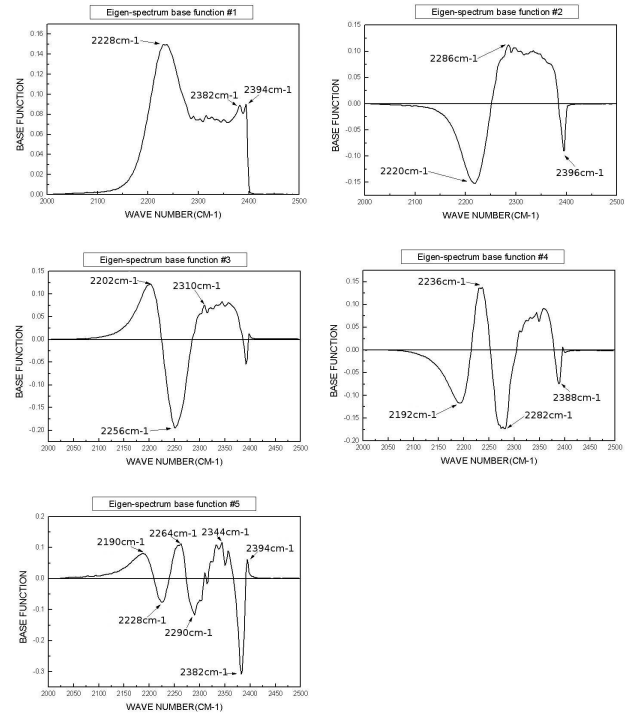


Figure 3. First 5 infrared region eigenspectrum ($2000\text{cm}^{-1} - 2500\text{cm}^{-1}$) with the wavenumber selected for the reduction of dimensionality.

neurons has been tested. The results are shown in table 3.

Table 3. Errors for the hot gas temperature retrieval using a MLP with 17 inputs and different architectures.

Hidden neurons	MSE Train	MSE Test	Mean error per profile (K)	Mean error hottest cell (K)
10	0.00238	0.00403	17.13	10.14
20	0.00123	0.00126	8.94	6.52
10x10	0.00117	0.00119	8.52	7.00
40	0.00066	0.00085	6.65	6.86
30	0.00068	0.00079	6.58	5.39

This second approach improve the results obtained using all the spectra wavelengths with a relative error of 0.64% in the hottest cell, a mean temperature error of 6.58 K and a relative error of length of 2.9% for the best case. It is interesting to look at the difference between this approach and the first one for architecture with 30 and 40 hidden layers. In first approach (see table 1) the MSE train error tends to decrease to 0 however the MSE test error tends to level off or grow which could imply that it is producing overfitting. Instead of this, in the second approach the errors tend to decrease and reach the convergence together without over-

fitting. This could be interpreted as if the large amount of data used would be acting as noise (also called the curse of dimensionality) for the retrieval, and the reduction we have done seems to reduce also this kind of noise.

4 Conclusions

Retrieval of temperature profile inside a hot gas cloud from the energy spectrum is a problem where neural networks can contribute in a efficient way to solve it. Classical physical-statistical techniques present the disadvantage of their low speed. In addition, they depend on the initial conditions and iterative inversion algorithms also need a high speed module to solve the direct model. On the other hand, neural networks present their speed as the most important advantage in this field. Once the MLP has been trained the inversion method is almost instantaneous.

CASIMIR has been used to calculate a data set at moderated spectral resolution (4 cm^{-1}) because conditions on the spatial resolution of the temperature profile are not very severe. However, an improvement of this spatial resolution will demand an improvement in the spectral resolution. This better resolution results in an energy spectrum with a very high number of wavenumbers. This high number of entries is a serious drawback to the use of neural networks in this framework.

In this work, a combination of MLP and a feature selection has been used with the purpose of approximate the temperature profile of the hot gas and to allow working with high spectral resolution.

The results presented in section 3.2 show the capability of neural networks to retrieve the temperature profile in a hot cloud composed by CO_2 and H_2O simulating a combustion environment. Retrieval is performed over the infrared emission spectrum with relative errors below 1% in the hottest cell (temperatures between 540–1140 K), giving satisfactory results. Reference values in literature show that a temperature measurement with an error of 0.5–1% is considered an accurate measurement in the range of 1200–2000 K. The DT025 sensor is a commercial "intrusive" thermocouple with a maximum error of 0.5%. This sensor is considered to be "extremely accurate". The digital image system [12] also has relative errors no greater than 1% in the range 1280–1690 °C. And systems based in diode laser absorption sensors have errors around 4% [1]. Furthermore, the results presented in section 3.3 equalize or even improve the previous ones using only the 7% of the spectrum data as input, and resolving the problem of dimensionality. Thus, a spectral feature selection using principal component analysis technique seems to be a good solution to get the variability of the spectrum when a perturbation in the temperature profile or length happened in order to reconstruct the temperature profile using neural networks.

References

- [1] *Proc. of the Combustion Institute*, 28:407–413, 2000.
- [2] F. Aires. Remote sensing from the infrared atmospheric sounding interferometer instrument. *J. Geophys. Res.*, 107(D22):ACH6–1–15, 2002.
- [3] F. Aires, A. Chédin, N. A. Scott, and W. B. Rossow. A regularized neural net approach for retrieval of atmospheric and surface temperatures with the iasi instrument. *Journal of Applied Meteorology*, 41:144–159, 2001.
- [4] S. Briz, A. J. de Castro, J. M. Aranda, J. Meléndez, and F. López. Reduction of false alarm rate in automatic forest fire infrared surveillance systems. *Remote Sens. Environ.*, 86:19–29, 2003.
- [5] G. Cybenko. Approximation by superposition of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2:303–314, 1989.
- [6] E. García-Cuesta. *CASIMIR: Cálculos Atmosféricos y Simulación de la Transmisión en el Infrarrojo*. University Carlos III L/PFC 01781, Madrid (in Spanish), 2003.
- [7] K. Hornik, M. Stinchcombe, and H. White. Multilayer feed-forward networks are universal approximators. *Neural Networks*, 2:359–366, 1989.
- [8] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Educ. Psychol.*, 24:417–441. 498–520, 1933.
- [9] H. L. Huang and P. Antonelli. Application of principal component analysis to high-resolution infrared measurement compression and retrieval. *J. Clim. Appl. Meteorol.*, 40:365–388, 2001.
- [10] I. T. Jolliffe. *Principal Component Analysis (2nd Ed.)*. Springer Series in Statistics Springer-Verlag, New York, 2002.
- [11] L. H. Liu and J. Jiang. Inverse radiation problem for reconstruction of temperature profile in axisymmetric free flames. *J. Quant. Spectrosc. Radiat. Transfer*, 70:207–215, 2001.
- [12] G. Lu, Y. Yan, and M. Colechin. A digital imaging based multifunctional flame monitoring system. *IEEE T. Instrum. Meas.*, 53:1152–1158, 2004.
- [13] N. J. McCormick. Inverse radiative transfer problems: a review. *Nuclear Science and Engineering*, 112(3):185–198, 1992.
- [14] K. Pearson. On lines and planes of closest fit to systems of points in space. *Phil. Mag.*, 2(6):559–572, 1901.
- [15] C. Romero, K. S. Li, X., and R. Rossow. Spectrometer-based combustion monitoring for flame stoichiometry and temperature control. *Appl. Therm. Eng.*, 25:659–676, 2005.
- [16] L. S. Rothman. The hitran molecular spectroscopic database: edition of 2000 including updates through 2001. *J. Quant. Spectrosc. Radiat. Transfer*, 2003.
- [17] D. Rumelhart, G. Hinton, and R. J. Williams. *Learning Internal Representations by Error Propagation*. MIT Press, Cambridge, 1986.
- [18] M. Thakur, A. Vyas, and C. Shakher. Measurement of temperature and temperature profile of an axisymmetric gaseous flames using lau phase interferometer with linear gratings. *Opt. Laser Eng.*, 36:373–380, 2001.
- [19] H. William, A. Saul, T. William, and P. Brian. *Numerical Recipes in C: The Art of Scientific Computing (2nd Ed.)*. Cambridge University Press, Cambridge, 1992.